MEDIOGRID System in Meteorological and Environment Applications

Mihaela Ordean
Research Department
iQuest Technologies
Motilor 6-8 Cluj-Napoca
Romania
MOrdean@iquestint.com

Cornelia Melenti
Computer Science Department
Technical University of Cluj-Napoca
Baritiu 26-28 Cluj-Napoca
Romania
Cornelia.Melenti@cs.utcluj.ro

Dorian Gorgan
Computer Science Department
Technical University of Cluj-Napoca
Baritiu 26-28 Cluj-Napoca
Romania
Dorian.Gorgan@cs.utcluj.ro

Abstract – **The MEDIOGRID system aims to accomplish a pilot program to process the images acquired in real time from meteorological and resource satellites by using grid computing technology, in order to extract the meteorological and environment parameters that characterize the atmospheric and terrestrial state. The immediate system application consists in early prognostication of wood fires and water floods.**

This paper intends to present the scope and proposed objectives of the system.

I. INTRODUCTION

The development of industry and generally of the human settlements has affected dramatically the ecosystems for the last decades. The level of noxious, pollution, carbon dioxide from atmosphere have generated atypical behaviour of the natural and meteorological phenomena – floods, fires, storms, rainfalls, droughts, strong winds. The supervising of such critical events requires a huge volume of surveillance data and extremely powerful real time computation.

Moreover the processing the data in real time involves huge hardware and software resources. The necessity of finding some global or regional solutions is as much obvious as the predictions of meteorologists for the next few years are extremely warning.

One of the research and experimental directions actually imposed by the related technology is the collection and the processing of the satellite images in real time by efficient computing systems like grid systems.

The instant images capture the status of a certain geographical area, and from the sequence of fragmented images the processing extracts the semantic information (like objects and their topology) and induces the dynamics and regional tendencies. The appropriate software applications should support the access of the specialists to huge data and complex operations for rapid analyzing and short term predictions.

The grid computing concept has been developed as a solution for the lack of computation resources in some scientific applications. First implementations were created to set in common some universities resources with their research laboratories in the scope of increasing the available computing power. The development of the microelectronics technology has determined the improvement of the computing and storing power and, for a while, the supercomputer concept was not on the top. The development of network technologies, of parallel and distributed computing concepts and moreover the characteristics of the human society (specializing and cooperation) have brought again in attention the concept of the grid computing but this time not only in the sphere of research but also in the field of real applications (e.g. financial markets, business, etc).

The first and the second sections of the paper present an introduction onto the grid computing technology and the processing of the environmental and meteorological images. Additionally, the second section presents the grid related works in Europe and Romania, and as well the satellite images based computing.

The scope and the objectives of the MEDIOGRID system are detailed in section III. Section IV highlights the implementation issues (i.e. architecture and virtual hosting environment) and the performance requirements of the system. The last section summarises the conclusions.

II. BACKGROUND

The grid systems are appropriate in the satellites based Earth observing applications, especially for requiring huge computing power.

The new MSG satellite (European meteorological geostationary satellite) is operational from 1st of March 2004, and follows the proximal launch of the first satellite of EPS series (European Polar System). Unlike previous systems, the increasing of data volume is with a factor of 30 (300 GB daily) and the computing system should process data in real time. For instance, 4 GB must be processed in less than a quarter of an hour (a new set of images is sent by the satellite at each 15 minutes, in twelve spectral bands). Therefore the grid systems could be an efficient and competitive solution [21].

The traditional supercomputer is transformed by this way into a parallel and very powerful computing network. The high speed network connects computing and storing resources or other equipments like radio telescope, satellite connections, etc and presents them as a unique computer – a virtual supercomputer. The requests for such a supercomputer are flexibility, security, easy access and operation.

One of the goal of conceiving such a virtual supercomputer is the achieving of the computing speed of petaflops. Nowadays, the most rapid supercomputer is IBM's ASCII White developed for the US Energy Department, with the computing speed of 12.3 teraflops. UK and US have implemented together the DTF project (Distributed Terascale Facility) to achieve the computing speed of 13,6 teraflops. The Royal Institute of Melbourne, Australia, has created a consortium VPAC, which can run scientific, engineering or commercial applications on a 128 processors system.

The first grid computing oriented applications are Globus [8], Grid Datafarm, GridBus [21], SimGrid, implemented in Linux and distributed as open sources.

In Europe there are more than 40 projects, in progress or already completed, in the domain of grid computing, but just a few of them are in the topics of environment and meteorology. For instance, we could mention the projects "Developing the basis for monitoring, modelling and predicting Space Weather" and "Cloud archive user service" [26]. The most of these projects are developed under the umbrella of the EU-DataGrid programme [22], the created consortiums being operational at the level of the origin country (CERN – Switzerland, ESA/ESRIN, INFN – Italy, PPARK - England, CNRS - France, FOM/NIKHEF - Nederland) or EGEE (Enabling Grids for E-science in Europe [9]) in the FP6 /IST programme.

In Romania, might be cited the initiative of Atomics Physics Institute, Magurele Platforme, which creates a consortium by MAGGI project (MAGurele Grid Initiative) [14], and also the project of West University of Timisoara COMPGRID. We mention also the RoGrid initiative of ICI [13]. None of these projects concerns on the environmental and meteorological research.

The MEDIOGRID project aims to implement a software platform based on parallel and distributed processing of the images acquired in real time from meteorological and resource satellites, and to experiment a pilot system supporting the dynamic analysis of ecological and social systems. The immediate applications of the MEDIOGRID system concerns on the early prognostication of the wood fires and the water floods.

The MEDIOGRID project is first Romanian proposal for a grid system based application in the study of the environmental and meteorological evolution. The pilot system is developed through a consortium of the Technical University of Cluj-Napoca, the National Meteorological Agency, and the Babes-Bolyai University.

III. SCOPE AND OBJECTIVES

The MEDIOGRID project purpose is to accomplish a pilot program which processes the images acquired in real time from meteorological and resource satellites. The already acquired satellite images are primary processed and sent to a grid system which divides them in geographical sub areas. The grid infrastructure processes data in a parallel and distributed manner and returns the performed and reassembled images of the initial geographical areas. The results allow the specialists to make interpretation and prediction on the environment phenomena of interest.

The project aims both general and particular objectives concerning the dynamics of the environmental and meteorological systems.

The project objectives consist mainly in finding, experimentation and implementation of the scientific and technical solutions for the concrete case of applications within the domain of studying the meteo and environment dynamics:

- Identify the environmental and meteorological information in the satellite images
- Define and experiment the classification techniques to split images
- Semanitical analysis of satellite images
- Parallel and distributed image processing
- Develop grid infrastructures for huge data processing
- Model the environmental and meteorological data in order to process them over the grid infrastructure
- Develop the software platform kernel (i.e. MEDIOGRID) for parallel and distributed processing
- Implement and test of the pilot system based on MEDIOGRID kernel
- Build and test particular applications for the study of the environmental and meteorological dynamics
- Define the methodology for the development of applications based on MEDIOGRID software platform

IV. IMPLEMENTATION ISSUES AND PERFORMACE REQUIREMENTS

The images acquired by satellite are primary processed and sent to the grid system, which divides them into geographical sub areas (see figure 1). Data is distributed and computed in parallel over the grid and the processed and reassembled images for the initial geographical areas are returned. The results allow the specialists to perform nowcasting (interpretations and predictions of environmental studied phenomena).

IV.1. System architecture

The MEDIOGRID system architecture (Fig. 1) contains the following base components:

- *Images acquisition system*: receives and stores satellite data;
- *Primary data processing system*: composes images and data correspondent to some spectral bands and sends them to the application server;
- *Application server*: communicates with the primary data processing system, divides the images, reassembles and processes final results;
- *GRID server*: administrates grid infrastructure, receives data from the application server and sends it to the grid network in order to be computed, receives the results and returns them to the application server;
- *Grid clients*: perform the basic processes on data specific for a particular geographical sub-area resulted from segmentation.

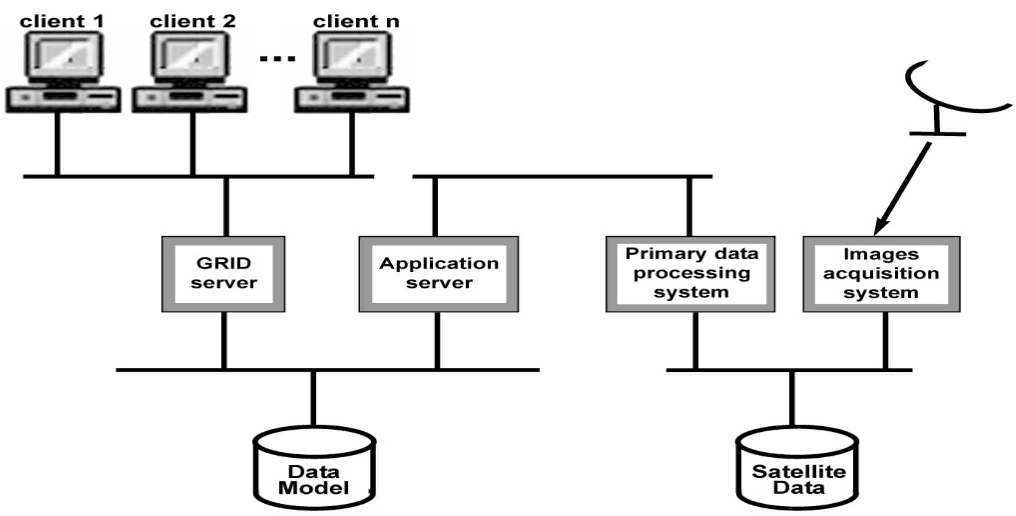

Fig 1. The MEDIOGRID system architecture

The transmission of primary data in order to be processed is accomplished in the first phase through an FTP client-server application in batch mode by using university network during application development and in inline mode during application testing. The image files and correspondent parameters, sent to the grid infrastructure, are in the range of hundreds of MB per session (approximating 200 MB once at 15 minutes). It is just the pre-processed data (spectral bands and interest parameters) instead of initial huge raw data (1.5 GB per hour).

IV.2. MEDIOGRID functional levels

Functional levels of the MEDIOGRID system are illustrated in figure 2. The MEDIOGRID platform layers the Grid Infrastructure that solves hardware requirements such as the information clusters. The VO (virtual organization) OGSA (Open Grid Service Architecture) [27] structure, used by the system as grid infrastructure software, will be one of virtual hosting environment VHE (figure 3). Some other platform components are data storing servers which will memorize the images obtained by the mean of acquisition system.

The next platform level provides primary services which assure either images acquisition or base services for grid infrastructure as well as acquired data processing. Among the services provided by VHE we can mention: data scheduling, resource co-allocation, resource reservation, administration and control functions.

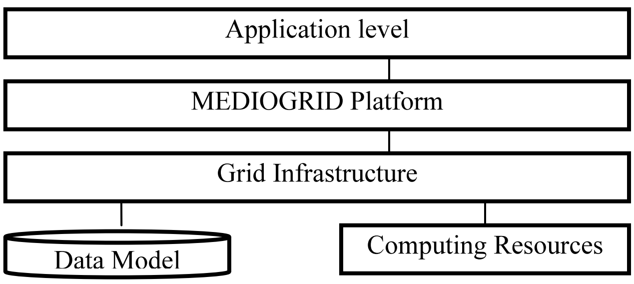

Fig.2. MEDIOGRID functional levels.

Images processing consist of images segmentation, pattern recognition, clouds shape determination, interpolation, isobars tracing. These are some of the parallel computings assured by the MEDIOGRID platform.

Application level solves the overall problems of the application like: data semantical analysis, environmental and meteorological parameters determination, sub-areas recomposing and whole image interpretation.

IV.3. Hosting environment

As we previously mentioned, the system will use a VO OGSA architecture, more precisely a virtual hosting environment VHE [23] (figure 3).

In order to create transient services and to discover and determine the properties of available services for a VO, an application can use the following OGSA interfaces: *Factory, Registry, GridService* and *HandleMap*. The interfaces can be used to create a variety of VO service structures. The client accesses the virtual hosting environment through the mentioned interfaces. For this purpose it should be created one or more “higher-level” factories that delegate the creation request to lower-level factories. Similarly, are going to be created higher-level registry that knows about higher-level factories and the service instances that they have created, as well as any VO-specific policies that govern the use of VO services. Clients can then use the VO registry to find factories and the VO services, and use their handlers to access them. Here, the registry handle is used as a global unique identifier in order to access the VO services.

The grid service mechanism will integrates both the geographically distributed resources and that of the local IT infrastructures. In all cases a collection of registered grid services can support functional capabilities delivering QoS interactions across distributed resource pool.

The application exploits these services for distributed resource management across heterogeneous platforms with local and remote transparency and locally optimized flows.

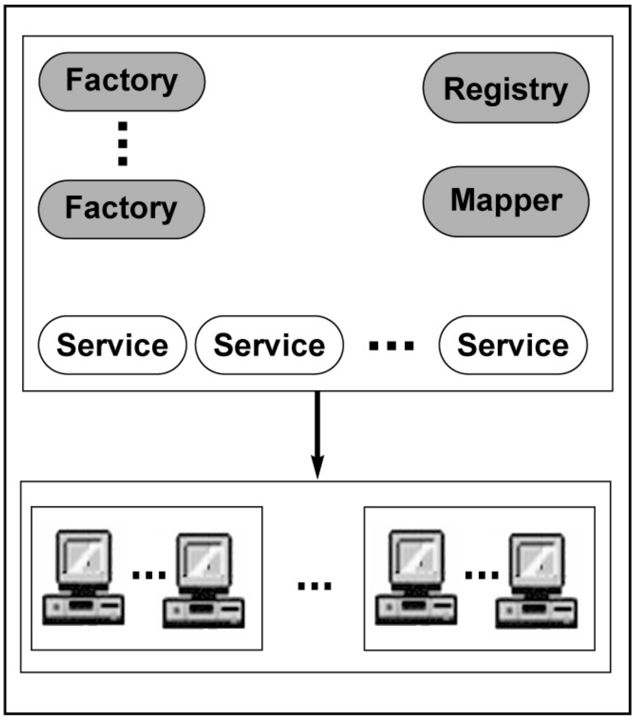

Fig. 3. Virtual hosting environment

IV.4. Performance requirements

Technical development grid infrastructure solutions are oriented toward the satisfaction of the following performance requests:

- Grid infrastructure complexity and transparency by using pattern based programming
- Load minimization by processes parallelization
- Performance improvement by object based parallelism
- Faul tollerance by data replication
- Component based and hierarchical software development: reused components, base software platform, domain level oriented (satellite images processing and recognition, semantical relations), application level and device level
- The specialists access data and operations in order to accomplish complex tasks for environmental and meteorological results analysis.

V. CONCLUSIONS

The MEDIOGRID project covers 3 years, first year achieves and experiments the grid infrastructure, data analysis and processing techniques. The second year develops the Core Software Platform consisting of basic components that support the image segmentation and the parallel and distributed processing of data, followed by the experiments of the pilot processing system based on the grid infrastructure. The last year develops and experiments a framework specific application for the dynamical analysis of the meteorological and environmental systems.

VI. REFERENCES

1. P. Busetta, P. Lucchese, M. Pistore, P. Traverso - Supporting Composition of Distributed Business Processes, Workshop ASTRO (Automated Composition of Distributed Business Processes), ITC-IRST and University of Trento
2. Mario Cannataro, Domenico Talia - Semantics and Knowledge Grids: Building the Next-Generation Grid, IEEE Intelligent Systems , January 2004
3. Junwei Cao, Stephen A. Jarvis, Subhash Saini, Graham R. Nudd - GridFlow: Workflow Management for Grid Computing, 3st International Symposium on Cluster Computing and the Grid, May 2003
4. Nigel Davies, Adrian Friday, Oliver Storz - Exploring the Grid's Potential for Ubiquitous Computing, IEEE Pervasive Computing, april, 2004
5. Ian Foster, Adriana Iamnitchi - On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing, International Workshop "The Potential of GRID and Distributed Computing Activities in Romania", Bucharest, 20-21 April 2002
6. Yolanda Gil, Ewa Deelman, Jim Blythe, Carl Kesselman, Hongsuda Tangmunarunkit - Artificial Intelligence and Grids: Workflow Planning and Beyond, IEEE Intelligent Systems, January 2004
7. Bob Jones - Grids - the next killer application, Terena Networking Conference, Rhodes, June 2004
8. Rick Stevens, Michael E. Papka, Terry Disz - Prototyping the Workspaces of the Future, IEEE Internet Computing, July 2003
9. EGEE - Enabling Grids for E-science in Europe http://www.eu-egee.org
10. Grid Project http://www.cern.ch
11. GIS and Mapping Software http://www.esri.com
12. IEEE Collection http://www.computer.org/
13. Romanian grid forum http://www.grid.ro/
14. Magurele MAGI Platform http://venus.nipne.ro/maggi/
15. Large-scale research projects in grid computing http://www.grid.org/
16. GRID Infoware http://www.gridcomputing.com/
17. GridWeaver project http://www.gridweaver.org
18. Grid forum http://www.ggf.org/
19. Globus toolkit http://www.globus.org
20. Network e-Commerce http://www.et-diamond.net/
21. GRID Testbed Web Portals http://www.gridbus.org/gridscape/
22. DataGRID project http://web.datagrid.cnr.it
23. EnterTheGrid project http://enterthegrid.com/
24. UK National e-Science Centre http://www.nesc.ac.uk/
25. Distributed Systems http://edamok.itc.it/
26. European Projects http://www.cordis.lu/
27. Open Grid Service Architecture http://www.gridforum.org/ogsi-wg/drafts/ogsa_draft2.9_2002-06-22.pdf